

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application Transmittal

Assistant Commissioner of Patents
Washington, D.C. 20231

Sir:

Transmitted herewith for filing is the **Patent Application** of:

Inventor: J. Blakely, et al

For: Method and system for Dynamic Creation of Mixed Language HTML Content Through MT

Enclosed are:

- ☒ Two (2) sheets of drawings.
- ☒ An assignment of the invention to International Business Machines Corporation, Armonk, New York 10504.
- ☐ A certified copy of a _____ application.
- ☐ An associate power of attorney.
- ☒ Declaration and Power of Attorney for Patent Application

The filing fee has been calculated as shown below:

	(Col. 1)	(Col. 2)	Other Than Small Entity	
For:	No. Filed	No. Extra	Rate	Fee
Basic Fee				\$690.00
Total Claims	9-20 =	0	x \$18.00=	\$.00
Indep. Claims	3 -3 =	0	x \$78.00=	\$.00
<input type="checkbox"/> Multiple Dependent Claim Presented			\$260.00	\$.00
Subtotal				\$690.00
Surcharge-Late Filing Fee or Oath or Declaration			\$130.00	\$.00
TOTAL				\$690.00

Deposit Account Authorization

- ☒ Please charge Deposit Account No. 09-0461 in the amount of \$690. A duplicate copy of this sheet is enclosed.
- ☒ The Commissioner is hereby authorized to charge payment of the following fees associated with this communication or credit any overpayment to Deposit Account No. 09-0461. A duplicate copy of this sheet is enclosed.
- ☒ Any additional filing fees required under 37 C.F.R. §1.16.
- ☒ Any patent application processing fees under 37 C.F.R. §1.17.

Date: May 23, 2000

Respectfully submitted,

By

A. Bruce Clay
A. Bruce Clay
Attorney of Record
Registration No. 32,121
IBM Corporation
Intellectual Property Law
Dept. T81/Bldg. 062
P.O. Box 12195
Research Triangle Park, NC 27709
Telephone: 919-254-6717
Fax: 919-254-4330

EXPRESS MAIL CERTIFICATE

Express Mail Label
Number: BH751370945US
Date: May 23, 2000

I hereby certify that I am depositing the enclosed or attached paper with the U.S. Postal Service "Express Mail Post Office to Addressee" service on the above date, addressed to the Assistant Commissioner of Patents, Washington, D.C. 20231

Catherine M. Robbins
Catherine M. Robbins

Docket No. RSW9-1999-0104

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: J. Blakely, et al

Serial No.: Group No.:

Filed: Herewith

For: Method and System for Dynamic Creation of Mixed Language HTML Content Through MT

Assistant Commissioner of Patents
Washington, D.C. 20231

EXPRESS MAIL CERTIFICATE

"Express Mail" label number: EH751370945US

Date of Deposit: May 23, 2000

WE REQUEST THE DATE OF DEPOSIT AS THE DATE FILED.

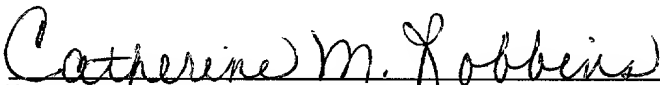
I hereby certify that the following **attached** correspondence comprising:

1. Patent Application Transmittal (In duplicate)
2. Declaration and Power of Attorney
3. Recordation and Assignment
4. Patent Application (12 Pages)
5. Formal Drawings (2)
6. Business Reply Post Card

is being deposited with the United States Postal Service Express Mail Post Office to Addressee service under CFR 1.10 on the date indicated above and is addressed to:

**Assistant Commissioner of Patents
Washington, D.C. 20231**

Catherine M. Robbins
(Name of person mailing paper or fee)


(Signature of person mailing paper or fee)

EXPRESS MAIL LABEL NO.: EH751370945US DATE OF DEPOSIT: 5/23/00

I hereby certify that this paper and fee are being deposited with the United States Postal Service Express Mail Post Office to Addressee service under 37 CFR §1.10 on the date indicated above and is addressed to the Assistant Commissioner of Patents, Washington, D.C. 20231.

Catherine M. Robbins

Catherine M. Robbins

INVENTORS: Jason Blakely; Dennis King and Richard Redpath

METHOD AND SYSTEM FOR DYNAMIC CREATION OF MIXED LANGUAGE HTML CONTENT THROUGH MT

Field of the Invention

This invention relates in general to computer software, and in particular to a method and system for dynamic creation of mixed language HTML content through Machine Translation.

Background of the Invention

Machine Translation (MT) is an emerging technology for the translation of text from one language to another. A number of vendors such as Systran, L&H, Transparent Languages, etc. provide Machine Translation software. On-The-Fly (OTF) MT is a unique approach to MT. OTF MT has been introduced in WebSphere Application Server (WAS) 3.0 (trademark of IBM Corp.) via the IBM HTTP Server (IHS) for static HTML and CGIs and for dynamic servlet content via the

application server. OTF MT allows for automatic MT initiated by configuration, user preference, control data or other reasons none of which require direct human intervention once configured . The mechanism for OTF MT in WAS 3.0 is based on configuration settings to indicate that MT is allowed and by a value in an Accept-Language field of the request-header for an HTTP request. This is a unique approach to OTF MT. The Accept-Language field is set by the browser based on user preferences. This approach provides a user preferred yet administrator authorized solution and is very valuable. These OTF MT techniques are associated with an entire HTML page or file of content. All of the textual language in the page is the target of translation. As content creators and host systems become more global in nature, a need is rapidly emerging whereby it is required that language content within the same page be mixed. No known solutions currently provide the ability to dynamically (OTF) translate different parts of an HTML page from any language into any other.

Summary of the Invention

The present invention describes a new technique using standard HTML tags and attribute declarations to specify the need for mixed language content for translation as well as the language to translate it to. Thus mixed language is possible within the same page of content. Once specified, the actual MT can be dynamically initiated.

In one aspect of the invention, a method and system is provided for determining a target language for automatic programmatic translation of text in a first language. An author first creates text in the first language (most likely his native language). An HTML 'lang' attribute is used to determine at least one target language that is different from the first language. The first language is then automatic programmatically translated into at least one target language using the 'lang' attribute as a key for machine translation. In addition, any appropriate method such as Language Guessing

may be used to determine the first language. A mixed language content can be created by using a plurality of target languages for translation into from the first language..

These and other features, aspects, and advantages of the present invention will become better understood with reference to the following description, appended claims, and accompanying drawings.

Brief Description of the Drawings

For a more complete understanding of the present invention and for further advantages thereof, reference is now made to the following Detailed Description taken in conjunction with the accompanying drawings, in which:

Figure 1 is a pictorial representation of a data processing system which may be utilized to implement a method and system of the present invention; and

Figure 2 is a schematic flow diagram of the present invention.

Detailed Description of the Invention

Referring to Figure 1, there is depicted a graphical representation of a data processing system 8, which may be utilized to implement the present invention. As may be seen, data processing system 8 may include a plurality of networks, such as Local Area Networks (LAN) 10 and 32, each of which preferably includes a plurality of individual computers 12 and 30, respectively. Of course, those skilled in the art will appreciate that a plurality of Intelligent Work Stations (IWS) coupled to

a host processor may be utilized for each such network. Each said network may also consist of a plurality of processors coupled via a communications medium, such as shared memory, shared storage, or an interconnection network. As is common in such data processing systems, each individual computer may be coupled to a storage device 14 and/or a printer/output device 16 and may be provided with a pointing device such as a mouse 17.

The data processing system 8 may also include multiple mainframe computers, such as mainframe computer 18, which may be preferably coupled to LAN 10 by means of communications link 22. The mainframe computer 18 may also be coupled to a storage device 20 which may serve as remote storage for LAN 10. Similarly, LAN 10 may be coupled via communications link 24 through a sub-system control unit/communications controller 26 and communications link 34 to a gateway server 28. The gateway server 28 is preferably an IWS which serves to link LAN 32 to LAN 10.

With respect to LAN 32 and LAN 10, a plurality of documents or resource objects may be stored within storage device 20 and controlled by mainframe computer 18, as resource manager or library service for the resource objects thus stored. Of course, those skilled in the art will appreciate that mainframe computer 18 may be located a great geographic distance from LAN 10 and similarly, LAN 10 may be located a substantial distance from LAN 32. For example, LAN 32 may be located in California while LAN 10 may be located within North Carolina and mainframe computer 18 may be located in New York.

Software program code which employs the present invention is typically stored in the memory of a storage device 14 of a stand alone workstation or LAN server from which a developer may access the code for distribution purposes, the software program code may be embodied on any

of a variety of known media for use with a data processing system such as a diskette or CD-ROM or may be distributed to users from a memory of one computer system over a network of some type to other computer systems for use by users of such other systems. Such techniques and methods for embodying software code on media and/or distributing software code are well-known and will not be further discussed herein.

A solution to the problem of how to create mixed translation of content within one document is based on a new use of the 'lang' attribute on standard HTML tags. The HTML 4.0 specification defines the 'lang' attribute as being used for an element's attribute value and content. Browsers use this information to assist in proper rendering of content associated with treatment of ligatures, hyphenations and quotations. Search engines and other applications may use this field for spell checking or creating indices. The implication of this common usage is that the content is in the language indicated by the 'lang' attribute and language specific rules for its rendering apply.

The present invention uses the 'lang' attribute value as a determining factor in the MT decision and allows multiple MT decisions within a page, thus enabling the creation of mixed language content documents. Prior to the content in a response being served, appropriate MT would be initiated based on this value. In addition to its current use, a new use of this attribute would be defined as the target language of MT. Now an HTML author or a program that dynamically creates content can do so in their language of choice yet the 'lang' attribute for an element would define the languages the various pieces of content should be translated to. Known techniques, such as, for example, Language Guessing could be used to determine the source language. Normal use of the same exact 'lang' attribute would still be applicable in terms of its use to make target rendering decisions since the MT of the content would occur prior to the use of the 'lang' attribute in rendering decisions. This solution could be implemented in, for example, the IBM HTTP Server and/or the

IBM WebSphere Application Server (trademarks of IBM Corp.) or any other system that supports or processes HTML.

As an example, consider a kiosk set up in a public location such as an airport. The kiosk is intended for people at the airport to get information about the location of various services available inside the airport terminal, such as, where a first aid station is located. Since an airport is likely to have travelers that speak many languages, the kiosk application must present directions in these many languages at the same time to be useful. The present invention could be applied to the system to present a screen that contains mixed language content in one screen, thus allowing one screen to contain directions for the same location in many languages at the same time. As a result of the present invention, the creator of the kiosk system can develop the screen in one language only and have it programmatically rendered in any number of mixed content views thus saving time, money, disk storage space, etc. As defined herein programmatic means actions that are initiated by a computer program. There is no requirement for the user to set any preference or take any action. For the present invention the user is not involved in the MT decision.

Referring to Figure 2, a scenario for use of the present invention would be:

The initial screen for this kiosk would contain international icons (i.e. pictures) indicating the various items it is capable of providing directions for. For this example, there are icons for restrooms, first aid, telephones, and police. These icons by definition are language neutral and understood by speakers of a number of native languages. The user would select the icon for which directions are desired (for example, first aid). The application was developed to display directions in four languages (i.e. French, Spanish, German and Italian). The user would be presented with an interface that contains four sets of directions equally spaced on the screen. Each set of directions

would be in a different language but identifying the same location. As a result of the present invention, the author of the directions would only have to create the directions in one language. If the author speaks English, the directions could be created in English only, even though English is not one of the languages presented by the kiosk.

5

An application flow to implement the above scenario could be as follows:

A kiosk could be comprised of a Server 200 and a Browser 202 (the Server 200 and Browser 202 could be located on the same machine or on separate interconnected machines. The server could include a Kiosk Application 204 and a storage device 206. An MT System 208 could be co-located with the Server 200 or with a separate interconnected machine.

10

Through conventional means, the Kiosk Application 204 creates a screen that contains international icons (i.e. pictures), and presents it to the user on Browser 202, as indicated by reference numeral 500. Based on which icon (i.e., per this example, first aid) is selected by the user, through any appropriate conventional means such as a touch screen as indicated by reference numeral 501, the Kiosk Application 204 retrieves from storage device 206 (or creates) directions in one particular language (the one language the directions were created in, in this case English), as indicated by reference numeral 502.

15

20

Through conventional means, the Kiosk Application 204 creates an HTML screen at 503 with the previously retrieved or created directions repeatedly included a number of times equal to that of the number of languages supported by this application (in this example, four). Since the directions were created in English, all four sets of directions will be in English at this point.

The Kiosk Application 204 would set at 504 the 'lang' attribute of each set of directions to the desired language (i.e., to French, Spanish, German or Italian).

The Kiosk Application 204 would then send the HTML page at 505 to the browser. However, just prior to the HTML being sent to the browser, the MT system 208 would be invoked to determine if MT is required based on the contents of the 'lang' attribute in each piece of content in the HTML page.

MT would be invoked for each piece of content that has a 'lang' attribute different than that of the associated piece of content. Each set of content would be translated based on its unique language requirement, as identified in its associated 'lang' attribute. In this example, the four sets of English will be translated at 507, sequentially, into the four other languages needed for display at the Browser 202. The newly translated content is then sent at 508 to the Browser 202 for display to the user in French, Spanish, German and Italian.

Thus, the present invention provides the benefits of being able to author text in only one language. The text can be translated on-the-fly into any other languages and can then be displayed in any desired mix of languages on a single page.

Although the present invention has been described with respect to a specific preferred embodiment thereof, various changes and modifications may be suggested to one skilled in the art and it is intended that the present invention encompass such changes and modifications as fall within the scope of the appended claims.

What is Claimed:

- 1 1. A method of determining a target language for automatic programmatic translation of text in
2 a first language, comprising the steps of:
3 creating text in the first language;
4 using an HTML 'lang' attribute to set at least one target language which is different from the
5 first language; and
6 automatically programmatically translating the first language into said at least one target
7 language with said 'lang' attribute as a key for machine translation.
- 1 2. The method of Claim 1, further comprising the step of:
2 using Language Guessing to determine the first language.
- 1 3. The method of Claim 1, wherein said at least one target language comprises a plurality of
2 languages resulting in translation into a mixed language content.
- 1 4. A system for determining a target language for automatic programmatic translation of text
2 in a first language, comprising:
3 creating text in the first language;

4 using an HTML 'lang' attribute to set at least one target language different from the first
5 language; and

6 automatically programmatically translating the first language into said at least one target
7 language with said 'lang' attribute as a key for machine translation.

1 5. The system of Claim 4, further comprising:

2 means for using Language Guessing to determine the first language.

1 6. The system of Claim 4, wherein said at least one target language comprises a plurality of
2 languages resulting in translation into a mixed language content.

1 7. A computer program product recorded on computer readable medium for determining a
2 target language for automatic programmatic translation of text in a first language, comprising:

3 computer readable means for creating text in the first language;

4 computer readable means for using an HTML 'lang' attribute to set at least one target
5 language different from the first language; and

6 computer readable means for automatically programmatically translating the first language
7 into said at least one target language.

1 8. The program product of Claim 7, further comprising:

2 computer readable means for using Language Guessing to determine the first language.

9. The program product of Claim 7, wherein said at least one target language comprises a plurality of languages resulting in translation into a mixed language content.

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2433	2434	2435	2436	2437	2438	2439	2440	2441	2442	2443
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Abstract

METHOD AND SYSTEM FOR DYNAMIC CREATION OF MIXED LANGUAGE HTML CONTENT THROUGH MT

A mixed translation of content is created in one document based on a new use of the 'lang' attribute on standard HTML tags. The HTML 4.0 specification defines the 'lang' attribute as being used for an element's attribute value and content. Browsers use this information to assist in proper rendering of content associated with treatment of ligatures, hyphenations and quotations. Search engines and other applications may use this field for spell checking or creating indices. The implication of this common usage is that the content is in the language indicated by the 'lang' attribute and language specific rules for its rendering apply

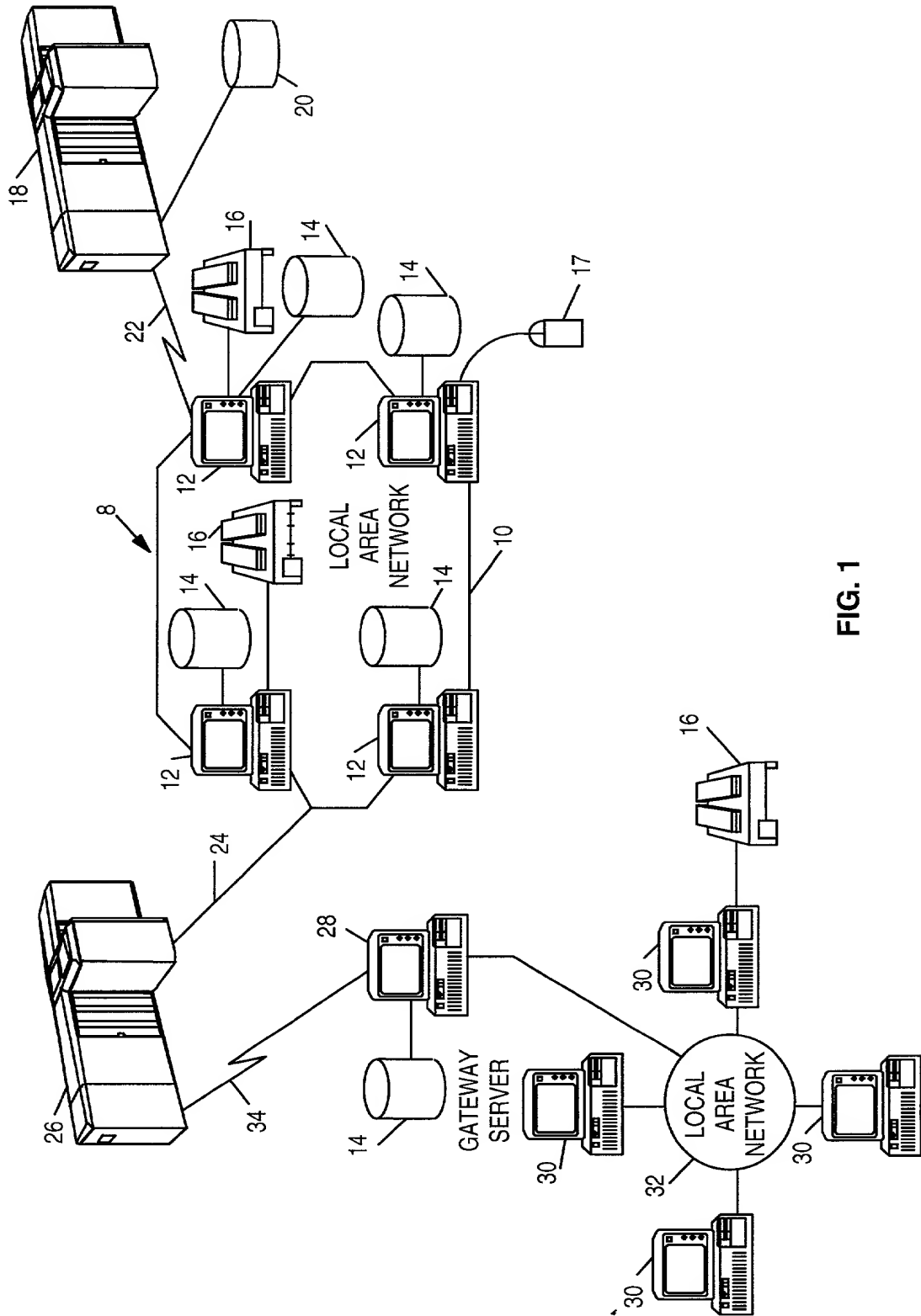


FIG. 1

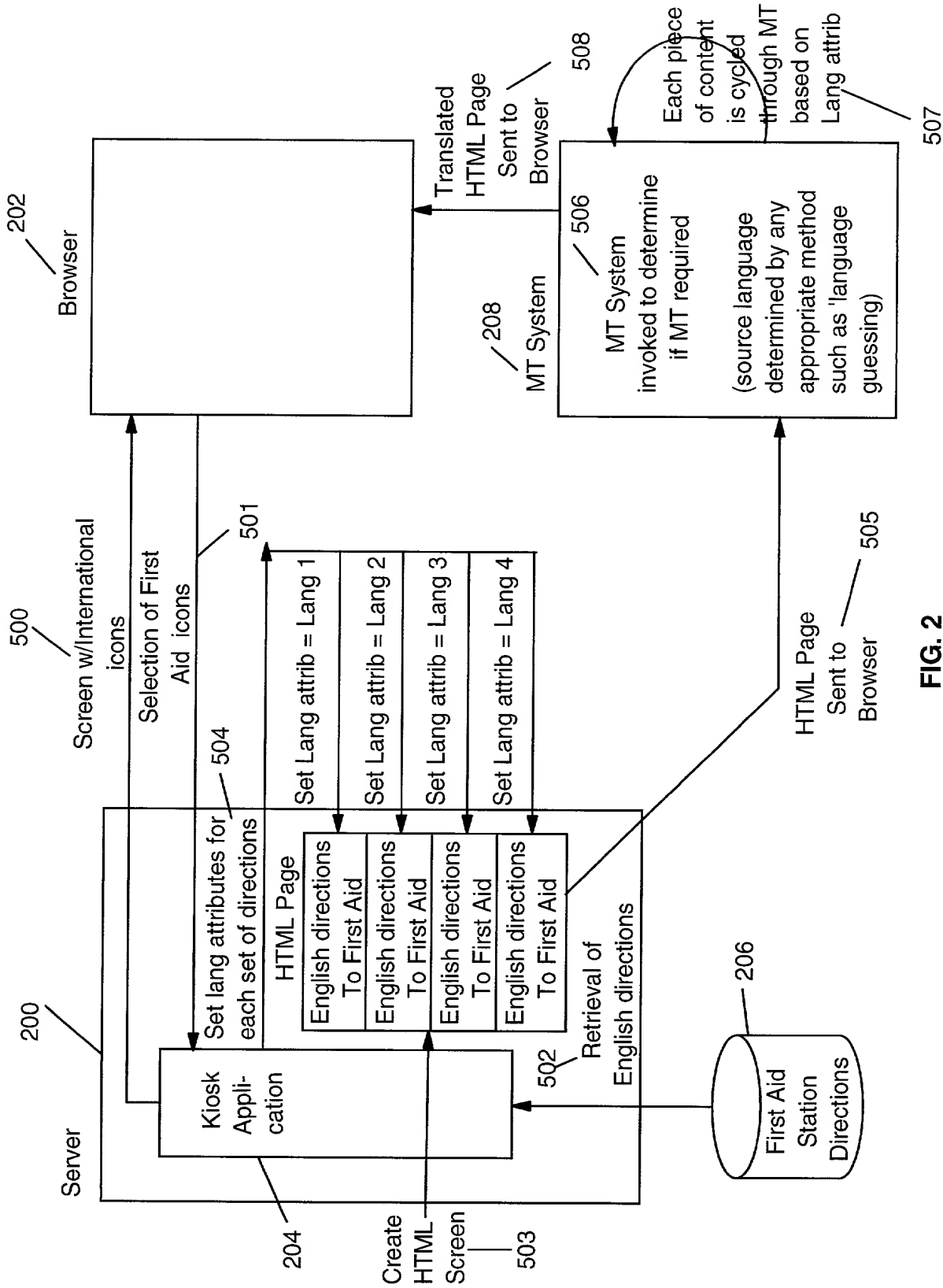


FIG. 2

Declaration and Power of Attorney for Patent Application

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

METHOD AND SYSTEM FOR DYNAMIC CREATION OF MIXED LANGUAGE HTML CONTENT THROUGH MT

the specification of which (check one)



is attached hereto.



was filed on _____ as Application Serial No. _____.

I hereby state that I have reviewed and understand the contents of the above- identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the patentability of this application in accordance with Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s):

Number

Country

Day/Month/Year

Priority Claimed

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose information material to the patentability of this application as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Prior U.S. Applications:

Serial No.

Filing Date

Status

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

As a named inventor, I hereby appoint the following attorneys and/or agents to prosecute this application and transact all business in the Patent and Trademark Office connected therewith:

A.B. Clay, Reg. No. 32,121; G. M. Doudnikoff, Reg. No. 32,847; E. H. Duffield, Reg. No. 25,970;
J. W. Herndon, Reg. No. 27,901; J. S. Ray-Yarletts, Reg. No. 39,808; Gerald R. Woods, Reg. No. 24,144

Send all correspondence to: A. Bruce Clay
IBM Corporation, Dept. T81/062
3039 Cornwallis Road
RTP, NC 27709
919-254-6717
FAX: 254-4330

- (1) Inventor: **Jason Y. Blakely**
- Signature:  Date 5/22/00
- Residence: 202 Wentbridge Road
Apex, North Carolina 27502
- Citizenship: USA
- Post Office Address: Same
- (2) Inventor: **Dennis D. King**
- Signature:  Date 5/22/2000
- Residence: 105 Saint Lenville Drive
Cary, North Carolina 27511
- Citizenship: USA
- Post Office Address: Same
- (3) Inventor: **Richard J. Redpath**
- Signature:  Date 19 May 2000
- Residence: 109 Blythewood Court
Cary, North Carolina 27513
- Citizenship: USA
- Post Office Address: Same